

# Representation of (Left) Ideal Regular Languages by Synchronizing Automata

Marina Maslennikova<sup>1</sup>, Emanuele Rodaro<sup>2</sup>

<sup>1</sup> Institute of Mathematics and Computer Science  
Ural Federal University, Ekaterinburg, Russia

<sup>2</sup> Centro de Matemática, Faculdade de Ciências  
Universidade do Porto, 4169-007 Porto, Portugal

maslennikova.marina@gmail.com, emanuele.rodaro@fc.up.pt

**Abstract.** We follow language theoretic approach to synchronizing automata and Černý’s conjecture initiated in a series of recent papers. We find a precise lower bound for the reset complexity of a principal ideal languages. Also we show a strict connection between principal left ideals and synchronizing automata. We characterize regular languages whose minimal deterministic finite automaton is synchronizing and possesses a reset word belonging to the recognized language.

**Keywords:** ideal language, synchronizing automaton, reset word, reset complexity, reset left regular decomposition, strongly connected automaton.

## Introduction

Let  $\mathcal{A} = \langle Q, \Sigma, \delta \rangle$  be a *deterministic finite automaton* (DFA), where  $Q$  is the *state set*,  $\Sigma$  stands for the *input alphabet*, and  $\delta : Q \times \Sigma \rightarrow Q$  is the totally defined *transition function* defining the action of the letters in  $\Sigma$  on  $Q$ . The function  $\delta$  is extended uniquely to a function  $Q \times \Sigma^* \rightarrow Q$ , where  $\Sigma^*$  stands for the free monoid over  $\Sigma$ . The latter function is still denoted by  $\delta$ . In the theory of formal languages the definition of a DFA usually includes the *initial state*  $q_0 \in Q$  and the set  $F \subseteq Q$  of *terminal states*. In this case a DFA is defined as a quintuple  $\mathcal{A} = \langle Q, \Sigma, \delta, q_0, F \rangle$ . We will use this definition when dealing with automata as devices for recognizing languages. A language  $L \subseteq \Sigma^*$  is said to be *recognized* (or *accepted*) by an automaton  $\mathcal{A} = \langle Q, \Sigma, \delta, q_0, F \rangle$  if  $L = \{w \in \Sigma^* \mid \delta(q_0, w) \in F\}$ , in this case we put  $L = L[\mathcal{A}]$ . We also use standard concepts of the theory of formal languages such as regular language, minimal automaton etc. [13]

A language  $I \subseteq \Sigma^*$  is called a *two-sided ideal* (or simply an *ideal*) if  $I$  is non-empty and  $\Sigma^* I \Sigma^* \subseteq I$ . A language  $I \subseteq \Sigma^*$  is called a *left* (respectively, *right*) *ideal* if  $I$  is non-empty and  $\Sigma^* I \subseteq I$  (respectively,  $I \Sigma^* \subseteq I$ ). In what follows we will consider only languages which are regular, thus we will drop the term “regular” and henceforth a given language will be implicitly a regular language. If it is said “ideal language” or simply “ideal”, it means that exactly a two-sided ideal language is considered, otherwise it will be explicitly mentioned which class of languages we are focusing on.

A DFA  $\mathcal{A} = \langle Q, \Sigma, \delta \rangle$  is called *synchronizing* if there exists a word  $w \in \Sigma^*$  whose action leaves the automaton in one particular state no matter at which state in  $Q$  it is applied, i.e.  $\delta(q, w) = \delta(q', w)$  for all  $q, q' \in Q$ . Any word with this property is said to be *reset* for the DFA  $\mathcal{A}$ . For the last 50 years synchronizing automata received a great deal of attention. For a brief introduction to the theory of synchronizing automata we refer the reader to the survey [19].

Recently in a series of papers [6, 9, 10, 17] a language theoretic (and descriptonal complexity) approach to the study of synchronizing automata has been developed. In the present paper we continue to study synchronizing automata from a language theoretic point of view and find a new approach to the Černý conjecture in this way. We denote by  $\text{Syn}(\mathcal{A})$  the language of reset words for a given synchronizing automaton  $\mathcal{A}$ . It is well known that  $\text{Syn}(\mathcal{A})$  is a regular language [19]. Furthermore, it is an ideal in  $\Sigma^*$ , i.e.  $\text{Syn}(\mathcal{A}) = \Sigma^* \text{Syn}(\mathcal{A}) \Sigma^*$ . On the other hand, every ideal language  $I$  serves as the language of reset words for some automaton. For instance, the minimal automaton recognizing  $I$  is synchronized by  $I$  [10]. Thus synchronizing automata can be considered as a special representation of ideal languages. The complexity of such a representation is measured by the *reset complexity*  $rc(I)$  which is the minimal possible number of states in a synchronizing automaton  $\mathcal{A}$  such that  $\text{Syn}(\mathcal{A}) = I$ . Every such automaton  $\mathcal{A}$  is called *minimal synchronizing automaton* (for brevity, MSA). Let  $sc(I)$  be the *state complexity* of  $I$ , i.e. the number of states in the minimal automaton recognizing  $I$ . Since the minimal automaton recognizing  $I$  has  $I$  as the language of reset words, we clearly have  $rc(I) \leq sc(I)$ . Moreover, there are ideals  $I_n$  for every  $n \geq 3$  such that  $rc(I_n) = n$  and  $sc(I_n) = 2^n - n$ , see [10]. So representation of an ideal language by means of one of its MSAs can be exponentially more succinct than its “traditional” representation via minimal automaton. However, no reasonable algorithm is known for computing an MSA for a given language. One of the obstacles is that MSA is not uniquely defined. Furthermore, the problem of checking, whether a given synchronizing automaton with at least five letters is an MSA for a given ideal language, has recently been shown to be **PSPACE**-complete [9].

Another source of motivation for studying representations of ideal languages by means of synchronizing automata comes from the famous Černý’s conjecture [3]. In 1964 Černý constructed for each  $n > 1$  a synchronizing  $n$ -state automaton  $\mathcal{C}_n$  whose shortest reset word has length  $(n - 1)^2$ . Later Černý conjectured that those automata represent the worst possible case, that is, every synchronizing automaton with  $n$  states possesses a reset word of length at most  $(n - 1)^2$ . Despite intensive efforts of researchers, this conjecture still remains open. One can restate easily the Černý conjecture in terms of reset complexity. Let  $\|I\|$  be the minimal length of words in an ideal language  $I$ . The Černý conjecture holds true if and only if  $rc(I) \geq \sqrt{\|I\|} + 1$  for every ideal  $I$ . The latter inequality would provide the desired quadratic upper bound on the length of the shortest reset word of a synchronizing automaton.

Thus, a deeper study of reset complexity may help to shed light on this long-standing conjecture. In this language theoretic approach to the Černý conjecture,

strongly connected synchronizing automata play an important role. Recall that a DFA is called *strongly connected* if for each pair of different states  $(p, q)$  there exists a word mapping  $p$  to  $q$ . It is well known that the Černý conjecture holds true whenever it holds true for strongly connected automata [20]. In this regard, an interesting question was posed in [6]. The question concerns the problem of finding a strongly connected synchronizing automaton whose set of reset words is equal to a given ideal language. Indeed, while the minimal automaton recognizing an ideal language  $I$  is always a synchronizing automaton with a unique *sink* state (i.e. a state fixed by all letters), finding examples of strongly connected synchronizing automata  $\mathcal{A}$  with  $\text{Syn}(\mathcal{A}) = I$  is a non-trivial task. In [17] it is proved that such strongly connected automaton always exists for an ideal over alphabet of size at least two. The construction itself is non-trivial and rather technical. Furthermore, the upper bound on the number of states of the associated strongly connected automaton is a double exponential. The approach of [17] has the extra advantage of detaching the Černý conjecture from the automata point of view. This is achieved by introducing a purely language theoretic notion of *reset left regular decomposition* of an ideal. This notion will be recalled in Section 1. Here we just focus on the connection between these decompositions and the Černý conjecture. Given an ideal  $I$ , the size of the smallest reset left regular decomposition of  $I$  is denoted by  $\text{rdc}(I)$ . This value can be viewed as the number of states of the smallest strongly connected synchronizing automaton  $\mathcal{A}$  with  $\text{Syn}(\mathcal{A}) = I$ . It is clear that  $\text{rc}(I) \leq \text{rdc}(I)$  and we have

**Theorem 1.** [16, Theorem 6] *Černý's conjecture holds if and only if for any ideal  $I$  we have  $\text{rdc}(I) \geq \sqrt{||I||} + 1$ .*

Therefore, the importance of the studies of issues like finding more effective constructions of reset left regular decompositions (or equivalently their associated automata) is evident.

Another interesting observation is the following. For each  $n \geq 3$  the corresponding MSA's for the aforementioned ideals  $I_n$  (with  $\text{rc}(I_n)$  and  $\text{sc}(I_n) = 2^n - n$ ) turned out to be strongly connected. Thus one may expect that there always exists a strongly connected MSA for an ideal language. However, in [5] it has been shown that a strongly connected MSA for a given ideal language does not always exist. Moreover, there are ideals  $J_n$  for every  $n \geq 3$  such that  $\text{rc}(J_n) = n + 1$  and  $\text{rdc}(J_n) = 2^n$ . Thus the smallest strongly connected automaton having a given ideal language  $I$  as the language of reset words may be exponentially larger than an MSA for  $I$ .

Recall that an ideal  $I$  is called *finitely generated* if  $I = \Sigma^* U \Sigma^*$  for some finite set  $U \subseteq \Sigma^*$ . Such languages have been viewed as languages of reset words of synchronizing automata in [14, 15]. Note that the aforementioned languages  $J_n$  are finitely generated ideals. In [6] it is considered the partial case of *principal* ideal languages, i.e. languages of the form  $\Sigma^* w \Sigma^*$ , for some  $w \in \Sigma^*$ . If  $|w|$  denotes the length of  $w \in \Sigma^*$ , then we have

**Theorem 2** ([6]). *For the language  $\Sigma^* w \Sigma^*$ , there is a strongly connected automaton  $\mathcal{B}$  with  $|w| + 1$  states, such that  $\text{Syn}(\mathcal{B}) = \Sigma^* w \Sigma^*$ . Such an automaton can be constructed in  $O(|w|^2)$  time.*

In the present paper we enforce the previous result by showing that the automaton  $\mathcal{B}$  from Theorem 2 is actually an MSA for a given language. More precisely, we prove that  $rdc(I) = rc(I) = \|I\| + 1$ , for every principal ideal language  $I$ . In particular, this result solves an open question posed in [6] regarding the size of the minimal strongly connected synchronizing automaton for which a given principal ideal language serves as the language of reset words. We show that *principal left ideals*, i.e. ideals of the form  $\Sigma^*w$  for some word  $w$ , play also a fundamental role in Černý’s conjecture. Indeed, we characterize strongly connected synchronizing automata via homomorphic images of automata belonging to a particular class  $\mathcal{L}(\Sigma)$  of automata. The class  $\mathcal{L}(\Sigma)$  is formed by all the trim automata  $\mathcal{A} = \langle Q, \Sigma, \delta, q_0, \{q_0\} \rangle$  such that  $L[\mathcal{A}] = w^{-1}\Sigma^*w$  for some word  $w \in \Sigma^*$ . In Section 2 we reduce Černý’s conjecture to the same conjecture for the quotients of automata from the class  $\mathcal{L}(\Sigma)$ . In view of this connection we study automata recognizing languages of the form  $w^{-1}\Sigma^*w$  for some  $w \in \Sigma^*$ . We provide a compact formula to calculate the syntactic complexity of a language  $I = w^{-1}\Sigma^*w$ . This value is defined just by the length of  $w$  and by the quantity of distinct prefixes, suffixes and factors in  $w$ . Another interesting feature of such languages concerns the construction of the minimal automaton  $\mathcal{A}_w$  recognizing the language  $w^{-1}\Sigma^*w$ . It turns out that  $w \in \text{Syn}(\mathcal{A}_w)$ . Thus, in this context, we have that a word of the language recognized by the automaton is also a reset word for this automaton. Hence it is quite natural to ask in which cases the minimal automaton recognizing a given regular language  $L$  is synchronized by some word from  $L$ . Here we answer this question by proving a criterion for the minimal automaton recognizing  $L$  to be synchronized by some word from  $L$ . We state this criterion in terms of the notion of a constant of  $L$  introduced by Schützenberger [18]. The notion of a constant is widely studied and finds applications in bioinformatics and coding theory [2, 8].

## 1 Preliminaries

Let  $\mathcal{A} = \langle Q, \Sigma, \delta, q_0, F \rangle$  be a deterministic finite automaton. The corresponding triple  $\langle Q, \Sigma, \delta \rangle$ , where the initial state and the set of final states are deliberately omitted, is called the *underlying semiautomaton* of  $\mathcal{A}$ . If the transition function  $\delta$  is clear from the context, we will write  $q.w$  instead of  $\delta(q, w)$  for  $q \in Q$  and  $w \in \Sigma^*$ . This notation extends naturally to any subset  $H \subseteq Q$  by putting  $H.w = \{\delta(q, w) \mid q \in H\}$ . A DFA  $\mathcal{A} = \langle Q, \Sigma, \delta, q_0, F \rangle$  is called *trim* whenever each state  $q \in Q$  is reachable from  $q_0$  and each state  $t \in F$  is reachable from some state  $q \in Q$ .

In our context a (automaton) *homomorphism*  $\varphi : \mathcal{A} \rightarrow \mathcal{B}$  between the DFAs  $\mathcal{A} = \langle Q, \Sigma, \delta \rangle$  and  $\mathcal{B} = \langle T, \Sigma, \xi \rangle$  is a map  $\varphi : Q \rightarrow T$  preserving the action of letters, i.e.  $\varphi(\delta(q, a)) = \xi(\varphi(q), a)$  for all  $a \in \Sigma$ . Note that  $\varphi(Q)$  identifies a sub-automaton of  $\mathcal{B}$  denoted by  $\varphi(\mathcal{A})$ , and we say that  $\varphi(\mathcal{A})$  is a *homomorphic image* of  $\mathcal{A}$ . A binary relation  $\rho \subseteq Q \times Q$  is a *congruence* for the automaton  $\mathcal{A} = \langle Q, \Sigma, \delta, q_0, F \rangle$  if  $(q_1, q_2) \in \rho$  implies  $(\delta(q_1, u), \delta(q_2, u)) \in \rho$  for all  $u \in \Sigma^*$ ,  $q_1, q_2 \in Q$ . The *quotient automaton* of a DFA  $\mathcal{A}$  with respect to

a congruence  $\rho$  is denoted by  $\mathcal{A}/\rho = \langle Q/\rho, \Sigma, \delta', [q_0], F/\rho \rangle$ , where  $[q]$  denotes the  $\rho$ -class containing  $q$ , and the transition function  $\delta' : Q/\rho \times \Sigma \rightarrow Q/\rho$  is defined by the rule  $\delta'([q], u) = [\delta(q, u)]$ , for all  $u \in \Sigma^*$ ,  $q \in Q$ . We denote by  $\text{Cong}(\mathcal{A})$  the set of all the congruences of the DFA  $\mathcal{A}$ , the *index* of a congruence  $\rho \in \text{Cong}(\mathcal{A})$  is the cardinality of the state set of  $\mathcal{A}/\rho$ . For any integer  $k$ , we use the symbol  $\text{Cong}_k(\mathcal{A})$  to denote the (possibly empty) set of congruences on  $\mathcal{A}$  of index  $k$ .

Denote the  $i$ -th letter of a word  $w \in \Sigma^+$  by  $w[i]$  and the prefix  $w[1]w[2] \dots w[i]$  by  $w[1..i]$ . For indices  $1 \leq i < j \leq |w|$  we use the notation  $w[i..j]$  to indicate the factor  $w[i]w[i+1] \dots w[j]$ . If  $1 \leq i < j$  then we put  $w[j..i] = \varepsilon$ . For  $u, w \in \Sigma^*$  we say that  $u$  is a *prefix*, (*suffix* or *factor*, respectively) of  $w$  if  $w = uu_2$  ( $w = u_1u$  or  $w = u_1uu_2$ , respectively) for some  $u_1, u_2 \in \Sigma^*$ . We also write  $u \leq_p w$  ( $u \leq_s w$  or  $u \leq_f w$ , respectively) if  $u$  is a prefix (suffix or a factor of  $w$ , respectively). We write  $u <_p w$  ( $u <_s w$  or  $u <_f w$ ) if  $u$  is a proper prefix (suffix or factor, respectively) of  $w$ . For a given language  $L \subseteq \Sigma^*$  and  $w \in \Sigma^*$  we put  $Lw = \{xw \mid x \in L\}$ ,  $wL = \{wx \mid x \in L\}$ . The *left* (*right*) *quotient* of  $L$  by a word  $w$  is the set  $w^{-1}L = \{v \in \Sigma^* : vw \in L\}$  ( $Lw^{-1} = \{v \in \Sigma^* : vw \in L\}$ ). We recall the following definition from [17]:

**Definition 1.** A reset left regular decomposition is a collection  $\{I_i\}_{i \in F}$  of disjoint left ideals  $I_i$  on  $\Sigma^*$ , for some finite set  $F$ , satisfying the following two conditions.

- i) For any  $a \in \Sigma$  and  $i \in F$ , there is an index  $j \in F$  such that  $I_i a \subseteq I_j$ .
- ii) Let  $I = \biguplus_{i \in F} I_i$ . For any  $u \in \Sigma^*$  if there is an  $i \in F$  such that  $Iu \subseteq I_i$ , then  $u \in I$ .

Denote by  $\mathbf{RLD}_\Sigma$  the class of the reset left regular decompositions over  $\Sigma$ . The notation  $\mathbf{SCSA}_\Sigma$  stands for the class of all strongly connected synchronizing automata over  $\Sigma$ . In [17] it has been shown that an ideal language  $I$  is strongly connected if and only if it has a reset left regular decomposition. The proof of this statement provides a bijection between the classes  $\mathbf{RLD}_\Sigma$  and  $\mathbf{SCSA}_\Sigma$ . This fact was stated in the following theorem.

**Theorem 3 (Theorem 4, [17]).** The map  $\mathcal{A} : \mathbf{RLD}_\Sigma \rightarrow \mathbf{SCSA}_\Sigma$  defined by the rule

$$\mathcal{A} : \{I_i\}_{i \in F} \mapsto \mathcal{A}(\{I_i\}_{i \in F}) = \langle \{I_i\}_{i \in F}, \Sigma, \eta \rangle$$

with  $\eta(I_i, a) = I_j$  for  $a \in \Sigma$  if and only if  $I_i a \subseteq I_j$  is a bijection with inverse given by  $\mathcal{I} : \mathbf{SCSA}_\Sigma \rightarrow \mathbf{RLD}_\Sigma$  defined by the rule

$$\mathcal{I} : \mathcal{B} = \langle Q, \Sigma, \delta \rangle \mapsto \{I_q\}_{q \in Q} = \{\{u \in \Sigma^* : \delta(Q, u) = q\}\}_{q \in Q}.$$

## 2 Lower bounds for the reset complexity of principal ideal languages

In this section we prove that  $\text{rdc}(I) = \text{rc}(I) \geq n + 1$  for a principal ideal language  $I = \Sigma^* w \Sigma^*$  with  $|w| = n$ . First we recall some auxiliary facts and

definitions from [14]. Let us consider an automaton  $\mathcal{A} = \langle Q, \Sigma, \delta \rangle$ . For a word  $u \in \Sigma^*$ , the *maximal fixed set*  $m(u)$  is the largest subset of  $Q$  fixed by  $u$ , i.e.  $m(u) \cdot u = m(u)$ . Note that  $m(u) = Q \cdot u^{k(u)}$  for some minimal integer  $k(u)$  and it is not difficult to see that  $k(u) \leq |Q| - |m(u)|$  (see [14, Lemma 2]). A synchronizing DFA  $\mathcal{A} = \langle Q, \Sigma, \delta \rangle$  is called *finitely generated* if the language  $\text{Syn}(\mathcal{A})$  is a finitely generated ideal. The following theorem is proved using the same technique of [14, Theorem 4], for the sake of completeness we present the proof in the appendix.

**Theorem 4.** *Let  $\mathcal{A} = \langle Q, \Sigma, \delta \rangle$  be a finitely generated synchronizing automaton with  $|Q| = n$ . Then for any word  $v \in \Sigma^+$  we have that either  $v^{k(v)} \in \text{Syn}(\mathcal{A})$ , or there is a word  $\tau$  with  $|\tau| \leq n - 1$ , such that  $v^{k(v)}\tau v^{k(v)} \in \text{Syn}(\mathcal{A})$ .*

We are now in position to prove the main theorem of this section.

**Theorem 5.** *Let  $I = \Sigma^* w \Sigma^*$  be a principal ideal language, then  $rc(I) = |w| + 1$ .*

*Proof.* Since in [10, Lemma 1] it has been shown that  $rc(I) = |w| + 1$  for  $w = a^n$ , we may assume  $|\Sigma| > 1$ . By Theorem 2 we have  $rc(I) \leq |w| + 1$ . Suppose, contrary to our claim, that there is a synchronizing automaton  $\mathcal{A} = \langle Q, \Sigma, \delta \rangle$  with  $|Q| = n \leq |w|$  for which  $I$  serves as the language of reset words. The equality  $|w| = 1$  implies that  $rc(I) = 2$ , so in what follows we assume that  $|w| > 1$ . Let  $a$  and  $b$  be the initial and final letter of  $w$  respectively. Denote by  $a^r$  the maximal prefix of  $w$  of the form  $a^l$ ,  $l \in \mathbb{N}$ , and by  $b^h$  the maximal suffix of  $w$  of the form  $b^l$ ,  $l \in \mathbb{N}$ . We consider the following cases.

**Case 1.** Assume  $a \neq b$ . Thus  $w$  can be factorized as  $w = a^r u b^h$  for some  $u \in \Sigma^*$ . Suppose first that  $u \in \Sigma^+$ . Let us take  $v = a^{|w|} b^{|w|}$ . By Theorem 4 we have two cases: either  $v^{k(v)} \in \text{Syn}(\mathcal{A}) = I$ , or there is a word  $\tau$  with  $|\tau| \leq n - 1 \leq |w| - 1$  such that  $v^{k(v)}\tau v^{k(v)} \in I$ .

Suppose that  $v^{k(v)} \in I$ . Thus  $w \leq_f v^{k(v)}$ , and since  $w$  can not be a factor of either  $a^{|w|}$  or  $b^{|w|}$ , it must be a factor of  $v$ . Since  $u \neq \varepsilon$  we have that  $u[1] \neq a$  and  $u[|u|] \neq b$  by the definition of  $a^r, b^h$ . Thus  $w$  is not a factor of  $v$ , a contradiction. Therefore, we can assume that  $v^{k(v)}\tau v^{k(v)} \in I$ , and so  $w \leq_f v^{k(v)}\tau v^{k(v)}$ . From the arguments above we have that  $w$  can not be a factor of  $v$  or  $v^{k(v)}$ , so we have  $w \leq_f v\tau v$ . Since  $w$  is not a factor of  $v$ ,  $w[1] = a \neq b$ ,  $w[|w|] = b \neq a$ , we obtain  $w \leq_f \tau$ . Hence  $|w| \leq |\tau| \leq |w| - 1$ , which is a contradiction.

Hence we may consider  $u = \varepsilon$ , and so  $w = a^r b^h$ . In [10, Lemma 1] it was shown that  $rc(I) = |w| + 1$  for  $w \in \{a^n, b^n\}$ . In the same paper it was obtained that  $rc(I) = |w| + 1$  for  $w = a^{n-1}b$ , thus we can assume that  $r \geq 1$  and  $h \geq 2$ . If  $r > 1$  we take  $v = a^{r-1}b^{h-1}$ . By Theorem 4 we have that either  $v^{k(v)} \in I$ , or  $v^{k(v)}\tau v^{k(v)} \in I$  for some word  $\tau$  with  $|\tau| \leq n - 1 \leq |w| - 1$ . Obviously,  $w = a^r b^h$  can not be a factor of  $v^{k(v)}$ . Therefore,  $w$  is a factor of  $v\tau v$ . Again using simple technique from combinatorics on words it is easy to see that  $w$  must be a factor of  $\tau$ . Hence we get  $|w| \leq |\tau| \leq |w| - 1$ , a contradiction. If  $r = 1$  we take  $v = ab^{h-1}$ . By Theorem 4 we have that either  $v^{k(v)} \in I$ , or  $v^{k(v)}\tau v^{k(v)} \in I$  for some word  $\tau$  with  $|\tau| \leq n - 1 \leq |w| - 1$ . The word  $w = ab^h$  is not a factor of  $v^{k(v)}$ , thus  $w \leq_f v\tau v$ . Note that  $h > 2$ , hence  $w$  must be a factor of  $\tau$ , which is again a contradiction.

**Case 2.** Assume  $a = b$ . If  $w \in \{a^n, b^n\}$  then  $rc(I) = |w| + 1$  [10, Lemma 1]. Therefore, we can assume that  $w = a^r u a^h$  for some  $u \in \Sigma^+$  with  $u[1] \neq a$ ,  $u[|u|] \neq a$ . In this case we apply Theorem 4 with  $v = b$  for some  $b \in \Sigma \setminus \{a\}$ . Providing the same arguments as above, it is easy to prove that  $w$  has to be a factor of a word  $\tau$  with  $|\tau| \leq |w| - 1$ , which again leads to the contradiction  $|w| \leq |\tau| \leq |w| - 1$ .  $\square$

Note that by Theorem 2 we have the equality  $rc(I) = rdc(I) = |w| + 1$ .

### 3 A lifting property for strongly connected synchronizing automata

The aim of this section is to prove that strongly connected synchronizing automata are all and only all the homomorphic images of automata from some particular class.

**Definition 2.** *The considered class  $\mathcal{L}(\Sigma)$  is formed by all the trim automata  $\mathcal{A} = \langle Q, \Sigma, \delta, q_0, \{q_0\} \rangle$  such that  $L[\mathcal{A}] = w^{-1} \Sigma^* w$  for some word  $w \in \Sigma^*$ .*

Here we reduce Cerný's conjecture to the same conjecture for the quotients of automata from the class  $\mathcal{L}(\Sigma)$ . We have the following proposition.

**Proposition 1.** *Let  $\mathcal{A} \in \mathcal{L}(\Sigma)$  with  $L[\mathcal{A}] = w^{-1} \Sigma^* w$ . Then  $\mathcal{A}$  is a strongly connected synchronizing automaton and  $w$  is a reset word for  $\mathcal{A}$ .*

*Proof.* Since  $\mathcal{A} = \langle Q, \Sigma, \delta, q_0, \{q_0\} \rangle$  is a trim DFA, for each  $q \in Q$  there is a word  $u \in \Sigma^*$  such that  $q_0 \cdot u = q$ . On the other hand,  $uw \in w^{-1} \Sigma^* w = L[\mathcal{A}]$ , thus we have  $q_0 = q_0 \cdot uw = q \cdot w$ . In this way, we obtain that  $q \cdot w = q_0$  for each  $q \in Q$ , i.e.  $w \in \text{Syn}(\mathcal{A})$ .

Now we prove that  $\mathcal{A}$  is a strongly connected DFA. Take two arbitrary states  $q_1, q_2 \in Q$ . Since  $\mathcal{A}$  is a trim DFA there is a word  $u$  such that  $q_0 \cdot u = q_2$ . Thus, since  $q_1 \cdot w = q_0$ , we have  $q_1 \cdot (wu) = q_0 \cdot u = q_2$ .  $\square$

Let  $w, u \in \Sigma^*$ , we denote by  $u \wedge_s w$  the maximal suffix of the word  $u$  that appears in  $w$  as a prefix. We have the following lemma (for the proof see appendix).

**Lemma 1.** *For any  $u, v, w \in \Sigma^*$ ,  $(uv) \wedge_s w = ((u \wedge_s w)v) \wedge_s w$ . Furthermore, for any  $v$  with  $|v| \geq w$ ,  $(uv) \wedge_s w = v \wedge_s w$ .*

Let  $\mathcal{A} = \langle Q, \Sigma, \delta, q_0, F \rangle$  be a DFA. For a state  $q \in Q$  we define the *right language* of  $q$   $L_q[\mathcal{A}] = \{u \in \Sigma^* \mid q \cdot u \in F\}$ . For  $p, q \in Q$  we say that  $p$  and  $q$  are *equivalent* if  $L_p[\mathcal{A}] = L_q[\mathcal{A}]$ . A DFA with a distinguished initial state and distinguished set of final states is minimal if it contains no (different) equivalent states and all states are reachable from the initial state. The automata from  $\mathcal{L}(\Sigma)$  recognize languages which are left quotients of the form  $w^{-1} \Sigma^* w$ . In fact these languages are recognized by automata with exactly  $|w| + 1$  states as it is shown in the following proposition.



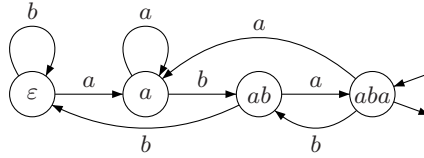
**Proposition 2.** Consider the automaton  $\mathcal{A}_w = \langle P(w), \Sigma, \xi, q_n, \{q_n\} \rangle$  where  $P(w) = \{q_0, \dots, q_n\}$  is the set of prefixes of the word  $w$  of length  $0 \leq i \leq |w| = n$ , and the transition function is defined by the rule  $\xi(q_i, a) = (q_i a) \wedge_s w$  for all  $a \in \Sigma$ ,  $q_i \in P(w)$ . The DFA  $\mathcal{A}_w$  is the minimal automaton recognizing the language

$$L[\mathcal{A}_w] = w^{-1} \Sigma^* w \quad (1)$$

*Proof.* By Lemma 1 it is straightforward to see that  $\xi(q_i, u) = (q_i u) \wedge_s w$  for all  $u \in \Sigma^*$ ,  $q_i \in Q$ . First we prove the equality (1). Let  $u \in \Sigma^*$  and  $\xi(q_n, u) = q_n$ . Hence  $w = q_n = (wu) \wedge_s w$ , i.e.  $wu \in \Sigma^* w$ . Conversely, if  $u \in w^{-1} \Sigma^* w$ , that is  $wu \in \Sigma^* w$ , then  $(wu) \wedge_s w = w = q_n$ . This implies that  $\xi(q_n, u) = q_n$ , i.e.  $u \in L[\mathcal{A}_w]$ .

We now consider the minimality issue. We verify that each state  $q_i \in P(w)$  is reachable from the initial state  $q_n$ . Indeed, let  $a$  be any letter from  $\Sigma$  different from  $w[1]$ . We have the equality  $\xi(q_n, a^n) = q_0$ . The word  $w[1..i]$  maps  $q_0$  to  $q_i$ , so we have  $\xi(q_n, a^n w[1..i]) = q_i$ . Now we take any  $q_i, q_j \in P(w)$  with  $i \neq j$ . Without loss of generality we can assume  $i < j$ . Consider the word  $u = w[j+1, n]$ . We have  $\xi(q_j, u) = q_n$  while  $\xi(q_i, u) \neq q_n$  since  $|q_i u| < |w|$ . Hence  $q_i, q_j$  are not equivalent. So the DFA  $\mathcal{A}_w$  is minimal.  $\square$

*Example 1.* Take  $w = aba$ ,  $\Sigma = \{a, b\}$ . The minimal automaton  $\mathcal{A}_w$  recognizing the language  $L = w^{-1} \Sigma^* w$  is shown in Fig. 1.



**Fig. 1.** Automaton  $\mathcal{A}_w$  for  $w = aba$

Let  $\mathcal{A} = \langle Q, \Sigma, \delta \rangle$  be a strongly connected synchronizing automaton. By Theorem 3 we can build for  $\mathcal{A}$  the associated reset left regular decomposition  $\mathcal{I}(\mathcal{A}) = \{I_i\}_{i \in Q}$  where  $\uplus_{i \in Q} I_i = I = \text{Syn}(\mathcal{A})$ . Take a word  $w \in I$  of minimum length. Let  $\sigma_w$  be a binary relation on  $I$  defined as follows. For  $u, v \in I$  we say that

$$(u, v) \in \sigma_w \text{ if and only if } u, v \in I_i \text{ for some } i \in Q \text{ and } u \wedge_s w = v \wedge_s w \quad (2)$$

We have the following lemma.

**Lemma 2.** Let  $\mathcal{A} = \langle Q, \Sigma, \delta \rangle$  be a strongly connected DFA, and  $\{I_i\}_{i \in Q}$  its associated reset left regular decomposition. The relation  $\sigma_w$  is a right congruence on  $I$ . Furthermore, each  $\sigma_w$ -class is a left ideal contained in  $I_i$  for some  $i \in Q$ .

*Proof.* See appendix.



Note that  $\mathcal{A}_w \in \mathcal{L}(\Sigma)$ . Now we are in position to state the main result of this section.

**Theorem 6.** *Let  $\mathcal{A} = \langle Q, \Sigma, \delta \rangle$  be a strongly connected synchronizing automaton. For any reset word  $w$  of minimum length, there is a DFA  $\mathcal{B} \in \mathcal{L}(\Sigma)$  with  $L[\mathcal{B}] = w^{-1}\Sigma^*w$  and*

$$\Sigma^*w\Sigma^* \subseteq \text{Syn}(\mathcal{B}) \subseteq \text{Syn}(\mathcal{A})$$

such that  $\mathcal{A}$  is a homomorphic image of  $\mathcal{B}$ .

*Proof.* See appendix.

Note that the previous theorem is constructive and we can effectively compute the lifted automaton  $\mathcal{B}$  of the statement. Moreover, the minimality of the length of the word  $w$  among the reset words is also necessary to ensure the fact that each equivalence class is a left ideal. We have the following corollary.

**Corollary 1.** *The class of strongly connected synchronizing automata are all and only all the homomorphic images of the class  $\mathcal{L}(\Sigma)$  formed by the trim automata  $\mathcal{A} = \langle Q, \Sigma, \delta, \{q_0\}, q_0 \rangle$  such that  $L[\mathcal{A}] = w^{-1}\Sigma^*w$  for some word  $w \in \Sigma^*$ .*

*Proof.* By Proposition 1 we have that any  $\mathcal{A} \in \mathcal{L}(\Sigma)$  is a strongly connected synchronizing automata, hence any homomorphic image  $\varphi(\mathcal{A})$  is also a strongly connected synchronizing automaton. On the other hand, by Theorem 6 any strongly connected synchronizing automaton is a homomorphic image of a DFA from  $\mathcal{L}(\Sigma)$ .  $\square$

Using Theorem 6 we can give another reformulation of Cerny's conjecture using the automata from  $\mathcal{L}(\Sigma)$ .

**Theorem 7.** *Cerny's conjecture holds if and only if for any  $\mathcal{B} \in \mathcal{L}(\Sigma)$  and  $\rho \in \text{Cong}_k(\mathcal{B})$  for all  $k < \sqrt{\|\text{Syn}(\mathcal{B})\|} + 1$  we have*

$$\|\text{Syn}(\mathcal{B}/\rho)\| < \|\text{Syn}(\mathcal{B})\|$$

*Proof.* See appendix.

## 4 Some properties of the automaton $\mathcal{A}_w$

In view of the results of the previous section, left quotients of principal left ideals seem to play a fundamental role in the Černý conjecture. In this regard we initiate a study of automata recognizing languages of the form  $w^{-1}\Sigma^*w$ . In this section we provide a compact formula to calculate the size of the syntactic semigroup of a language  $I = w^{-1}\Sigma^*w$ ,  $w \in \Sigma^*$ .

For a regular language  $L \subseteq \Sigma^*$  the *Myhill congruence* [12]  $\approx_L$  of  $L$  is defined as follows:

$$u \approx_L v \text{ if and only if } xuy \in L \Leftrightarrow xvy \in L \text{ for all } x, y \in \Sigma^*.$$

This congruence is also known as *the syntactic congruence* of  $L$ . The quotient semigroup  $\Sigma^+ / \approx_L$  of the relation  $\approx_L$  is called the *syntactic semigroup* of  $L$ . The syntactic semigroup of  $L$  is known to be isomorphic to the transition semigroup of the minimal DFA recognizing  $L$ . The *syntactic complexity*  $\sigma(L)$  of a regular language  $L$  is the cardinality of its syntactic semigroup. The notion of syntactic complexity is studied quite extensively: for a survey of this topic we refer the reader to [7]. Also the notion of the syntactic semigroup finds interesting application in the theory of synchronizing automata. Indeed, let  $I$  be an ideal language,  $\mathcal{S}$  the syntactic semigroup of  $I$  and  $\mathcal{S}(\mathcal{B})$  the transition semigroup of a synchronizing DFA  $\mathcal{B}$  for which  $I = \text{Syn}(\mathcal{B})$ . In [6] it has been shown that  $\mathcal{S}$  is a homomorphic image of  $\mathcal{S}(\mathcal{B})$ .

Recall that  $u \in \Sigma^+$  is an *inner factor* of  $w$  if there exist words  $x, y \in \Sigma^+$  such that  $w = xuy$ . Denote by  $\text{Fact}(w)$  the set of different inner factors of  $w$ , by  $\text{Suff}(w)$  the set of proper non-empty suffixes of  $w$  which do not appear in  $w$  as inner factors, by  $\text{Pref}(w)$  the set of proper non-empty prefixes of  $w$  which do not appear in  $w$  as suffixes or inner factors, by  $\text{Pref}_{\text{syn}}(w)$  the set of prefixes of  $w$  synchronizing  $\mathcal{A}_w$ . We have the following

**Proposition 3.** *Let  $I = w^{-1}\Sigma^*w$  for some  $w \in \Sigma^*$ . The syntactic complexity of  $I$  is equal to*

$$\sigma(I) = |w| + 1 + |\text{Pref}(w)| + |\text{Fact}(w)| + |\text{Suff}(w)| - |\text{Pref}_{\text{syn}}(w)|.$$

*Proof.* See appendix.

Note that by Proposition 3 we get an effective algorithm to calculate the syntactic complexity of the left quotient  $w^{-1}I$  by  $w$  of a principal left ideal  $I = \Sigma^*w$ .

By Proposition 1 the minimal automaton  $\mathcal{A}_w$  recognizing  $I = w^{-1}\Sigma^*w$  is strongly connected and  $w \in \text{Syn}(\mathcal{A}_w)$ . Further we show that  $\mathcal{A}_w$  is finitely generated. Recall that a reset word  $w$  for a given synchronizing DFA  $\mathcal{A}$  is called *minimal* if none of its proper prefixes nor suffixes belong to  $\text{Syn}(\mathcal{A})$ . Denote by  $\text{Syn}_{\text{min}}(\mathcal{A}_w)$  the set of all minimal reset words for a given synchronizing DFA  $\mathcal{A}_w$ .

**Proposition 4.** *For each  $w \in \Sigma^*$ ,  $\mathcal{A}_w$  is a finitely generated synchronizing automaton.*

*Proof.* In order to obtain the desired result we prove that the set  $\text{Syn}_{\text{min}}(\mathcal{A}_w)$  is finite. Take an arbitrary  $u \in \text{Syn}_{\text{min}}(\mathcal{A}_w)$ . If  $|u| > |w|$  then  $u$  is not minimal. Indeed, by the definition of the transition function of  $\mathcal{A}_w$  and by Lemma 1 we get, for all  $q_i \in P(w)$ ,  $q_i \cdot u = q_i u \wedge_s w = u \wedge_s w = u[2..|u|] \wedge_s w$  since  $|u| - 1 \geq |w|$ . Thus we have  $|u| \leq |w|$ . However, there is just finite amount of words of length at most  $|w|$ . Hence  $\mathcal{A}_w$  is a finitely generated synchronizing automaton.  $\square$

## 5 Representation of regular languages by synchronizing automata

In this section  $\mathcal{A}_L$  stands for the minimal DFA recognizing a regular language  $L$ . In some cases  $\mathcal{A}_L$  may have a unique *non-accepting* sink state  $s$ , i.e.  $s \notin F$ . It may turn out that  $\mathcal{A}_L$  is synchronizing and, therefore, each reset word brings the whole automaton to  $s$ . If this is not the case one may consider partial synchronization in the following sense. A DFA  $\mathcal{A} = \langle Q, \Sigma, \delta, q_0, F \rangle$  with a non-accepting sink state  $s$  is called *partially synchronizing* if there exists a word  $w \in \Sigma^*$  such that  $Q \cdot w = \{s, q\}$  for some state  $q \in Q$ . Any word with this property is said to be *partial reset* word for the DFA  $\mathcal{A}$ . And the set of all partial reset words for  $\mathcal{A}$  is denoted by  $\text{Syn}^{\text{par}}(\mathcal{A})$ .

Let  $L$  be a regular language. If  $L$  is an ideal language then  $\mathcal{A}_L$  is synchronizing and  $\text{Syn}(\mathcal{A}_L) = L$ . In Section 3 it has been shown that the minimal automaton recognizing the language  $w^{-1}\Sigma^*w$  is synchronizing and  $w$  is a reset word for this automaton. On the other hand,  $w \in w^{-1}\Sigma^*w$ . So in this case we have that the minimal automaton recognizing a given language  $L$  is synchronizing and some word from  $L$  is also a reset word for the automaton. In this regard the following interesting question arises. How to describe all regular languages  $L$  for which  $\mathcal{A}_L$  is synchronizing and  $L \cap \text{Syn}(\mathcal{A}_L) \neq \emptyset$ ? In this section we answer this question.

Let  $L \subseteq \Sigma^*$  be a regular language. A word  $w \in \Sigma^*$  is a *constant* for  $L$  if the implication

$$u_1wu_2 \in L, u_3wu_4 \in L \Rightarrow u_1wu_4 \in L$$

holds for all  $u_1, u_2, u_3, u_4 \in \Sigma^*$ . We denote the set of all constants of  $L$  by  $C(L)$ . Note that the set  $C(L)$  contains the ideal  $Z(L) = \{w \mid \Sigma^*w\Sigma^* \cap L = \emptyset\}$ . Constant words of a regular language  $L$  satisfy the following property, also stayed in [18].

**Lemma 3.** *Let  $L \subseteq \Sigma^*$  be a regular language and let  $\mathcal{A}_L$  be the minimal automaton recognizing  $L$  with set of states  $Q$ . If  $\mathcal{A}_L$  has a non-accepting sink state  $s$  then a word  $w \in \Sigma^*$  is a constant for  $L$  if and only if  $|Q \cdot w| \leq 2$ . If  $\mathcal{A}_L$  does not have a non-accepting sink state  $s$  then a word  $w \in \Sigma^*$  is a constant for  $L$  if and only if  $|Q \cdot w| = 1$ .*

By this lemma it follows that constants of a regular language  $L$  are described precisely via reset and partial reset words of the minimal automaton recognizing  $L$ . Let  $L \subseteq \Sigma^*$ , denote by  $\overline{L}$  the *complement* to  $L$ , that is  $\overline{L} = \Sigma^* \setminus L$ .

**Proposition 5.** *The automaton  $\mathcal{A}_L$  is synchronizing and  $L \cap \text{Syn}(\mathcal{A}_L) \neq \emptyset$  if and only if the following properties hold:*

- (i)  $C(L) \neq \emptyset$
- (ii)  $\overline{L}$  does not contain right ideals.

*Proof.* Consider the DFA  $\mathcal{A}_L = \langle Q, \Sigma, \delta, q_0, F \rangle$ . Assume that  $\mathcal{A}_L$  is synchronizing and the condition  $L \cap \text{Syn}(\mathcal{A}_L) \neq \emptyset$  holds. We take any  $w \in \overline{L} \cap \text{Syn}(\mathcal{A}_L)$ . By Lemma 3 we have  $w \in C(L)$ . Arguing by contradiction assume that  $\overline{L}$  contains a

right ideal. This means that there is a strongly connected component  $H \subseteq Q \setminus F$  without outgoing transitions leading to  $F$ . Thus, for all  $w \in \text{Syn}(\mathcal{A}_L)$ , we have  $H \cdot w \cap F = \emptyset$ , hence  $L \cap \text{Syn}(\mathcal{A}_L) = \emptyset$ , which is a contradiction.

Assume that properties (i) and (ii) hold. By property (ii)  $\mathcal{A}_L$  does not have a non-accepting sink state. Thus, by Lemma 3 each constant of  $L$  is a reset word for  $\mathcal{A}_L$ , and since  $C(L)$  is not empty,  $\mathcal{A}_L$  is synchronizing. Arguing by contradiction, assume that  $L \cap \text{Syn}(\mathcal{A}_L) = \emptyset$ , hence  $\text{Syn}(\mathcal{A}_L) \subseteq \overline{L}$ . However, the language  $\text{Syn}(\mathcal{A}_L)$  is a right ideal, a contradiction.  $\square$

The following proposition deals with the complementary case.

**Proposition 6.** *The automaton  $\mathcal{A}_L$  is synchronizing and  $L \cap \text{Syn}(\mathcal{A}_L) = \emptyset$  if and only if the following properties hold:*

- (i)  $Z(L) \neq \emptyset$
- (ii)  $\overline{L}$  contains a right ideal.

*Proof.* Consider the DFA  $\mathcal{A}_L = \langle Q, \Sigma, \delta, q_0, F \rangle$ . Assume that  $\mathcal{A}_L$  is synchronizing and the condition  $L \cap \text{Syn}(\mathcal{A}_L) = \emptyset$  holds. Arguing by contradiction assume that  $\overline{L}$  does not contain a right ideal. By Proposition 5 we get that  $L \cap \text{Syn}(\mathcal{A}_L) \neq \emptyset$ , which is a contradiction. So property (ii) holds. This property is equivalent to the existence of a strongly connected component  $H \subseteq Q \setminus F$ . By the minimality of  $\mathcal{A}_L$  we obtain  $|H| = 1$ , thus  $H$  contains just a non-accepting sink state. Since  $\mathcal{A}_L$  is synchronizing, each  $w \in \text{Syn}(\mathcal{A}_L)$  brings the whole DFA  $\mathcal{A}_L$  to  $s$ , hence  $Z(L) \neq \emptyset$ .

Conversely, assume that properties (i) and (ii) hold. Again, by property (ii) there is a non-accepting sink state in  $\mathcal{A}_L$ . Thus each  $w$  from  $Z(L)$  is a reset word for  $\mathcal{A}_L$ . Arguing by contradiction, assume that  $L \cap \text{Syn}(\mathcal{A}_L) \neq \emptyset$ . Thus by Proposition 5  $\overline{L}$  does not contain right ideals. Contradiction.  $\square$

Note that in order to check whether property (ii) in both of the previous propositions is satisfied, it is enough to check whether there is a strongly connected component in  $Q \setminus F$ . The latter can be implemented in time  $O(n \cdot |\Sigma|)$ , where  $n = |Q|$ . Note that some problems related two constants of languages are considered in [1]. In particular, the problem of deciding whether a partial 2-letter automaton is partially synchronizing is shown to be *NP*-complete (the action of the transition function on some states of a given automaton may be undefined). The notion of a partial synchronizing word from [1] is defined analogously to the notion of partial reset word here. Now we formally state the following **CONSTANT** problem:

- *Input:* a regular language  $L$  over  $\Sigma$ , presented via its minimal recognizing DFA  $\mathcal{A}_L$ .
- *Question:* is it true that  $C(L) \neq \emptyset$ ?

We can suppose that  $\mathcal{A}_L$  has a non-accepting sink state  $s$ , since otherwise the problem is equivalent to testing  $\mathcal{A}_L$  for synchronization in usual sense. First we prove the following

**Lemma 4.** *Let  $\mathcal{A}_L = \langle Q, \Sigma, \delta, q_0, F \rangle$  have a non-accepting sink state  $s$ . The set  $C(L)$  is not empty if and only if for each pair  $\{p, q\}$  of different states  $p, q \in Q$  there is a word  $u$  such  $\{p, q\} \cdot u \subseteq \{s, r\}$  for some  $r \in Q$ .*

*Proof.* Clearly, if  $C(L) \neq \emptyset$  the desired property holds by Lemma 3. Conversely, take any pair  $\{p, q\}$  of different states, then there is a word  $w_1 \in \Sigma^*$  such that  $\{p, q\} \cdot w_1 \subseteq \{s, r\}$  for some  $r \in Q$ . We clearly have  $|Q \cdot w_1| < |Q|$ . Consider now the set  $Q \cdot w_1$ . If  $|Q \cdot w_1| \leq 2$  then  $w_1 \in C(L)$ , so we are done. Otherwise, if  $|Q \cdot w_1| > 2$  then take again any two different states  $p', q' \in Q \cdot w_1$  such that  $p', q' \neq s$ . Hence there is a word  $w_2 \in \Sigma^*$  such that  $\{p', q'\} \cdot w_2 \subseteq \{s, r'\}$  for some  $r' \in Q$ . We have the inequality  $|Q \cdot w_1 w_2| < |Q \cdot w_1| < |Q|$ . Consider now the set  $Q \cdot w_1 w_2$ . If  $|Q \cdot w_1 w_2| \leq 2$  then  $w_1 w_2 \in C(L)$ , so we are done. Arguing by induction we get, through a finite number of steps as described above, a word  $w$  such that  $|Q \cdot w| \leq 2$ . That is  $w \in C(L)$ .  $\square$

Recall that for a given DFA  $\mathcal{A} = \langle Q, \Sigma, \delta, q_0, F \rangle$  the *power automaton*  $\mathcal{P}(\mathcal{A})$  is constructed as follows. Its state set  $\mathcal{Q}$  includes all non-empty subsets of  $Q$  and the transition function is a natural extension of  $\delta$  on the set  $\mathcal{Q} \times \Sigma$ . The latter function is still denoted by  $\delta$ . Denote by  $\mathcal{P}^{[2]}(\mathcal{A})$  the subautomaton of the power automaton  $\mathcal{P}(\mathcal{A})$  consisting only of 2-element and 1-element subsets of  $Q$ .

**Proposition 7.** *CONSTANT can be solved in time  $O(n^5 \cdot |\Sigma|)$ , where  $n = |Q|$ .*

*Proof.* We use Lemma 4 to establish nonemptiness of the set  $C(L)$ . First we build the corresponding automaton  $\mathcal{P}^{[2]}(\mathcal{A})$  that can be done in time  $O(n^2 \cdot |\Sigma|)$ . This automaton has  $\frac{n(n+1)}{2}$  states. Take any pair  $\{p, q\}$  of different states  $p, q \in Q$ ,  $p, q \neq s$ . Take any pair  $\{r, s\}$ ,  $r \neq s$ . We put  $L_{p,q,r,s} = \{w \mid \{p, q\} \cdot w = \{r, s\}\}$ ,  $L_{p,q,r} = \{w \mid \{p, q\} \cdot w = \{r\}\}$ ,  $L_{p,q,s} = \{w \mid \{p, q\} \cdot w = \{s\}\}$ . Nonemptiness of any of these three sets can be checked in time  $O(n^2 \cdot |\Sigma|)$  by a breadth first search in  $\mathcal{P}^{[2]}(\mathcal{A})$ . The latter may be done for all possible pairs  $\{p, q\}$  and  $\{r, s\}$  (in the worst case). Since there are  $\frac{n(n-1)^2}{2}$  possible choices for the pairs  $\{p, q\}$  and  $\{r, s\}$ , we get a cost of  $O(n^5 \cdot |\Sigma|)$ . Finally, we obtain that it takes  $O(n^5 \cdot |\Sigma|)$  time to solve CONSTANT.

**Remark.** Some partial results of the paper have been presented on the Third Russian Finnish Symposium on Discrete Mathematics RuFiDiM2014. The conference provided local proceedings (not indexed) in which we have presented an extended abstract of the communication without any proof.

## Acknowledgements

The first author acknowledges support from the Presidential Programme for young researchers, grant MK-3160.2014.1, from the Presidential Programme “Leading Scientific Schools of the Russian Federation”, project no. 5161.2014.1, and from the Russian Foundation for Basic Research, project no. 13-01-00852. The last author acknowledges support from the European Regional Development Fund through the programme COMPETE and by the Portuguese Government through the FCT – Fundação para a Ciência e a Tecnologia under the project PEst-C/MAT/UI0144/2013 as well as support from the FCT project SFRH/BPD/65428/2009.

## References

1. Berlinkov, M.V.: Testing for synchronization. CoRR abs/1401.2553 (2014)
2. Bonizzoni, P., De Felice, C., Zizza, Z.: The structure of reflexive regular splicing languages via Schützenberger constants. *Theor. Comp. Sci.*, vol. 334, pp. 71–98 (2005)
3. Černý, J.: Poznámka k homogénnym eksperimentom s konečnými automatami. *Mat.-Fyz. Cas. Slovensk. Akad. Vied.*, vol. 14, pp. 208–216, (1964) [in Slovak].
4. Eppstein, D.: Reset sequences for monotonic automata. *SIAM J. Comput.*, vol. 19, pp. 500–510, (1990)
5. Gusev, V.V., Maslennikova, M.I., Pribavkina, E.V.: Finitely generated ideal languages and synchronizing automata. In: J.Karhumäki, A.Lepistö, L.Zamboni (eds.), *WORDS 2013, LNCS*, vol. 8079, Springer, pp. 143–153 (2013)
6. Gusev, V.V., Maslennikova, M.I., Pribavkina, E.V.: Principal Ideal languages and synchronizing automata. In: V.Halava, J.Karhumäki, Yu. Matiyasevich (eds.), the Special Issue of the RuFiDiM 2012, *Fundamenta Informaticae*, vol. 132(1), pp. 95–108 (2014)
7. Holzer, M., König, B.: On deterministic finite automata and syntactic monoid size. *Theoret. Comput. Sci.* 327 (2004) P. 319–347
8. De Luca, A., Perrin, D., Restivo, A. and Termini, S.: Synchronization and simplification. *Discrete Math.*, vol. 27, pp. 297–308 (1979)
9. Maslennikova, M.: Complexity of checking whether two automata are synchronized by the same language. In: H.Jürgensen, J. Karhumäki, Al. Okhotin (eds.), *DCFS 2014, LNCS*, vol. 8614, Springer, pp. 306–317 (2014)
10. Maslennikova, M.I.: Reset Complexity of Ideal Languages. In: M. Bieleková (eds.), *SOFSEM 2012, Proc. Institute of Computer Science Academy of Sciences of the Czech Republic*, vol. II, pp. 33–44 (2012)
11. Margolis, S.W., Pin, J.-E., Volkov, M.V.: Words guaranteeing minimum image. *Int. J. Found. Comput. Sci.*, vol. 15(2), pp. 259–276 (2004)
12. Myhill, J.: Finite automata and representation of events. Wright Air Development Center Technical Report, 57624 (1957)
13. Perrin, D.: Finite automata. *Handbook of theoretical computer science*, J. van Leeuwen (eds.), Elsevier, B., pp. 1–57 (1990)
14. Pribavkina, E., Rodaro, E.: Synchronizing automata with finitely many minimal synchronizing words. *Information and Computation*, vol. 209(3), pp. 568–579 (2011)
15. Pribavkina, E.V., Rodaro, E.: Recognizing synchronizing automata with finitely many minimal synchronizing words is **PSPACE**-complete. In B. Löwe (eds.), *CiE 2011, LNCS*, vol. 6735, Springer-Verlag, Berlin-Heidelberg, pp. 230–238 (2011)
16. Reis, R., Rodaro, E.: Ideal regular languages and strongly connected synchronizing automata <http://www.dcc.fc.up.pt/dcc/Pubs/TReports/TR13/dcc-2013-01.pdf>
17. Reis, R., Rodaro, E.: Regular ideal languages and synchronizing automata. In: J. Karhumäki, A. Lepistö, L. Zamboni (eds.), *WORDS 2013, LNCS*, vol. 8079, pp. 205–216, Springer, Heidelberg (2013)
18. Schützenberger, M.P.: Sur certaines opérations de fermeture dans les langages rationnels. *Sympos. Math.*, vol.15, pp. 245–253 (1975) [in French].
19. Volkov, M.V.: Synchronizing automata and the Černý conjecture. In: C. Martín-Vide, F. Otto, H. Fernau (eds.), *Languages and Automata: Theory and Applications, LATA 2008, LNCS*, vol.5196. pp.11–27, Springer, Berlin (2008)

20. Volkov, M.V.: Synchronizing automata preserving a chain of partial orders. *Theor. Comp. Sci.*, vol. 410, pp. 3513–3519 (2009)



### Appendix

Let us consider an automaton  $\mathcal{A} = \langle Q, \Sigma, \delta \rangle$ . For a subset  $S \subseteq Q$ , denote by  $\text{Fix}(S)$  the set of words  $u \in \Sigma^*$  such that  $S \cdot u = S$ , and by  $\text{Syn}(S)$  the set of words  $\{u \in \Sigma^* \text{ such that } |S \cdot u| = 1\}$ . A subset  $S \subseteq Q$  is called *reachable* if  $Q \cdot u = S$  for some  $u \in \Sigma^*$ . We shall use the equality  $m(u^\ell) = m(u)$  for any  $\ell \geq 1$  and  $u \in \Sigma^*$  (see [14, Lemma 3]). The class of finitely generated synchronizing automata has the following combinatorial characterization.

**Theorem 8.** [14, Theorem 1] *A synchronizing automaton  $\mathcal{A} = \langle Q, \Sigma, \delta \rangle$  is finitely generated if and only if, for any reachable subset  $S \subseteq Q$  with  $1 < |S| < |Q|$  and for any  $u \in \text{Fix}(S)$  it holds that  $\text{Syn}(S) = \text{Syn}(m(u))$ .*

The *deficiency* of a word  $u \in \Sigma^*$  with respect to  $\mathcal{A}$  is the difference  $\text{df}(u) = |Q| - |Q \cdot u|$ . We make use of the following result from [11].

**Theorem 9.** *Given a synchronizing automaton  $\mathcal{A} = \langle Q, \Sigma, \delta \rangle$  and the words  $u, v \in \Sigma^+$  such that  $\text{df}(u) = \text{df}(v) = k > 1$ , there exists a word  $\tau$ , with  $|\tau| \leq k+1$ , such that  $\text{df}(u\tau v) > k$ .*

Now we are in position to prove Theorem 4.

**Theorem 4.** *Let  $\mathcal{A} = \langle Q, \Sigma, \delta \rangle$  be a finitely generated synchronizing automaton with  $|Q| = n$ . Then for every word  $v \in \Sigma^+$  we have that either the word  $v^{k(v)}$  is reset for  $\mathcal{A}$ , or there is a word  $\tau$  with  $|\tau| \leq n-1$ , such that  $v^{k(v)}\tau v^{k(v)}$  is a reset word for  $\mathcal{A}$ .*

*Proof.* Let us take an arbitrary word  $v \in \Sigma^+$ . If  $|m(v)| = 1$  then  $v^{k(v)} \in \text{Syn}(\mathcal{A})$ , so we are done. Now we may assume that  $|m(v)| > 1$ . We construct the following set

$$\text{REACH}(v) = \{S \subseteq m(v) \mid S = m(v) \cdot u, u \in \Sigma^*, |S| > 1\}$$

which is non-empty since  $m(v) \in \text{REACH}(v)$ . By Theorem 8 for any  $S \in \text{REACH}(v)$  it holds that

$$\text{Syn}(S) = \text{Syn}(m(v)). \quad (3)$$

Indeed, since  $S \subseteq m(v)$  we have  $v^\ell \in \text{Fix}(S)$  for some integer  $\ell > 1$ . On the other hand, synchronizing DFA  $\mathcal{A}$  is finitely generated. Thus applying Theorem 8 we get  $\text{Syn}(S) = \text{Syn}(m(v^\ell)) = \text{Syn}(m(v))$ . Now let  $H = Q \cdot v^{k(v)}u$  be an element of  $\text{REACH}(v)$  of minimal cardinality and let  $k' = n - |H| = \text{df}(v^{k(v)}u)$ . Since  $|H| > 1$  we have  $k' \leq n-2$ . Since  $\mathcal{A}$  is synchronizing we have by Theorem 9 that there is a word  $\tau$  with  $|\tau| \leq k' + 1 \leq n-1$  such that  $\text{df}(v^{k(v)}u\tau v^{k(v)}u) > k'$ , i.e.  $|Q \cdot v^{k(v)}u\tau v^{k(v)}u| < n - k' = |H|$ . We claim that the word  $v^{k(v)}u\tau v^{k(v)}uv^{k(v)}$  is reset for  $\mathcal{A}$ . Indeed,  $Q \cdot v^{k(v)}u\tau v^{k(v)}u \subseteq Q$ , thus  $Q \cdot v^{k(v)}u\tau v^{k(v)}uv^{k(v)} \subseteq Q \cdot v^{k(v)} = m(v)$ . So we obtain that  $Q \cdot v^{k(v)}u\tau v^{k(v)}uv^{k(v)}$  is an element of  $\text{REACH}(v)$  with

$$|Q \cdot v^{k(v)}u\tau v^{k(v)}uv^{k(v)}| \leq |Q \cdot v^{k(v)}u\tau v^{k(v)}u| < |H|.$$

Hence by the choice of  $H$  we get  $|Q \cdot v^{k(v)}u\tau v^{k(v)}uv^{k(v)}| = 1$ , that is the word  $v^{k(v)}u\tau v^{k(v)}uv^{k(v)}$  is reset for  $\mathcal{A}$ . In fact even the word  $v^{k(v)}u\tau v^{k(v)}$  is reset for

$\mathcal{A}$ . Indeed, consider the set  $S = Q \cdot v^{k(v)} u \tau v^{k(v)}$ . Let us assume that  $|S| > 1$ . In this case it holds that  $S \in \text{REACH}(v)$ , hence by (3) we obtain  $uv^{k(v)} \in \text{Syn}(S) = \text{Syn}(m(v))$ , so

$$1 = |m(v) \cdot uv^{k(v)}| = |Q \cdot v^{k(v)} uv^{k(v)}| = |H \cdot v^{k(v)}|.$$

But by the choice of  $H$  we have the inequality  $|H| > 1$ . Furthermore,  $H \subseteq m(v)$ . On the other hand,  $v$  acts as a permutation on  $m(v)$ . Therefore, we have  $|H \cdot v^{k(v)}| = |H| > 1$ , which is a contradiction and we get  $|S| = 1$ . Thus, since  $S = Q \cdot v^{k(v)} u \tau v^{k(v)} = H \cdot \tau v^{k(v)}$ , by (3) we obtain

$$\tau v^{k(v)} \in \text{Syn}(Q \cdot v^{k(v)} u) = \text{Syn}(H) = \text{Syn}(m(v)) = \text{Syn}(Q \cdot v^{k(v)}),$$

i.e.  $v^{k(v)} \tau v^{k(v)}$  is a reset word for  $\mathcal{A}$ .  $\square$

**Lemma 1.** *For any  $u, v, w \in \Sigma^*$ ,  $(uv) \wedge_s w = ((u \wedge_s w)v) \wedge_s w$ . Furthermore, for any  $v$  with  $|v| \geq w$ ,  $(uv) \wedge_s w = v \wedge_s w$ .*

*Proof.* Let  $t = (uv) \wedge_s w$ . If  $t <_s v$ , then it is easy to see that  $t = hv \wedge_s w$  where  $h$  is an arbitrary suffix of  $u$ . In particular, we have  $t = ((u \wedge_s w)v) \wedge_s w$ . Thus we can assume that  $t <_s uv$  and there is a non-empty word  $r \in \Sigma^+$  such that  $r \leq_s u$ ,  $r \leq_p w$  and  $t = rv$ . Hence  $r \leq_s (u \wedge_s w)$ . Since  $t$  is the maximal suffix of  $uv$  which is also a prefix of  $w$  and  $r \leq_s (u \wedge_s w) \leq_s u$  we get that  $t$  is also the maximal suffix of  $(u \wedge_s w)v$  which is also a prefix of  $w$ , i.e.  $t = ((u \wedge_s w)v) \wedge_s w$ . The last statement of the lemma follows trivially from the definition.  $\square$

**Lemma 2.** *Let  $\mathcal{A} = \langle Q, \Sigma, \delta \rangle$  be a strongly connected DFA, and  $\{I_i\}_{i \in Q}$  its associated reset left regular decomposition. The relation  $\sigma_w$  is a right congruence on  $I$ . Furthermore, each  $\sigma_w$ -class is a left ideal contained in  $I_i$  for some  $i \in Q$ .*

*Proof.* Clearly,  $\sigma_w$  is an equivalence relation on  $I$ . Let  $a \in \Sigma$  and  $(u, v) \in \sigma_w$ , i.e.  $u, v \in I_i$  for some  $i \in Q$  and  $u \wedge_s w = v \wedge_s w$ . By property *i*) of Definition 1 we have  $ua, va \in I_i a \subseteq I_j$  for some  $j \in Q$ . Furthermore, by Lemma 1 and  $u \wedge_s w = v \wedge_s w$  we have

$$(ua) \wedge_s w = ((u \wedge_s w)a) \wedge_s w = ((v \wedge_s w)a) \wedge_s w = (va) \wedge_s w$$

hence  $(ua, va) \in \sigma_w$ . Since the number of possible prefixes of  $w$  is finite, by the definition of  $\wedge_s$  we have that  $\sigma_w$  has finite index. Take any  $u \in I$ , denote by  $[u]$  the  $\sigma_w$ -class of  $u$ . Clearly,  $[u] \subseteq I_i$  for some  $i \in Q$ . Since  $w$  is a reset word for  $\mathcal{A}$  of minimum length, for each word  $u \in I$  we have  $|u| \geq |w|$ , thus for each  $v \in \Sigma^*$  we have  $(vu) \wedge_s w = u \wedge_s w$  by Lemma 1. Hence  $[u]$  is a left ideal in  $\Sigma^*$  contained in  $I_i$  for some  $i \in Q$ .  $\square$

**Theorem 6.** *Let  $\mathcal{A} = \langle Q, \Sigma, \delta \rangle$  be a strongly connected synchronizing automaton. For any reset word  $w$  of minimum length, there is a DFA  $\mathcal{B} \in \mathcal{L}(\Sigma)$  with  $L[\mathcal{B}] = w^{-1} \Sigma^* w$  and*

$$\Sigma^* w \Sigma^* \subseteq \text{Syn}(\mathcal{B}) \subseteq \text{Syn}(\mathcal{A})$$

*such that  $\mathcal{A}$  is a homomorphic image of  $\mathcal{B}$ .*

*Proof.* Let  $\mathcal{A} = \langle Q, \Sigma, \delta \rangle$  be a strongly connected synchronizing automaton. By Theorem 3 we can build for  $\mathcal{A}$  the associated reset left regular decomposition  $\mathcal{I}(\mathcal{A}) = \{I_i\}_{i \in Q}$  where  $\bigsqcup_{i \in Q} I_i = I = \text{Syn}(\mathcal{A})$ . Since  $w \in I$ , there is some  $j \in Q$  such that  $w \in I_j$  and thus  $\Sigma^*w \subseteq I_j$ . Let  $\sigma_w$  be a binary relation on  $I$  defined by (2). By Lemma 2 each  $\sigma_w$ -class is a left ideal contained in some  $I_i$  for some  $i \in Q$ .

Therefore  $\sigma_w$  induces a refinement  $\{J_t\}_{t \in T}$  of  $\{I_i\}_{i \in Q}$  for some set of indices  $T = \{v_0, \dots, v_m\}$ . Since  $\sigma_w$  is a right congruence, for any  $v_i \in T, a \in \Sigma$  we have  $J_{v_i}a \subseteq J_{v_j}$  for some  $T = \{v_0, \dots, v_m\}$ . Thus  $\Sigma$  defines an action  $\lambda$  on  $T$  defined by  $\lambda(v_i, a) = v_h$  where  $v_h$  is the unique index of  $T$  such that  $J_{v_i}a \subseteq J_{v_h}$ . Using a simple induction on the length of the words it is straightforward to check that the following condition holds

$$\lambda(v_i, u) = v_t, u \in \Sigma^* \text{ iff } J_{v_i}u \subseteq J_{v_t} \quad (4)$$

Note that  $\Sigma^*w$  is a  $\sigma_w$ -class belonging to  $\{J_t\}_{t \in T}$ , say  $\Sigma^*w = J_{v_0}$ . Therefore, consider the DFA  $\mathcal{B} = \langle H, \Sigma, \lambda, v_0, \{v_0\} \rangle$  where

$$H = \{v_j \in T : v_j = \lambda(v_0, u) \text{ for some } u \in \Sigma^*\}$$

and let us prove that  $\mathcal{B} \in \mathcal{L}(\Sigma)$  with  $L[\mathcal{B}] = w^{-1}\Sigma^*w$ . Since  $J_{v_h}w \subseteq \Sigma^*w = J_{v_0}$  for any  $v_h \in H$ , then by (4) we have  $\lambda(v_h, w) = v_0$ , so  $\mathcal{B}$  is a trim DFA. We now prove the equality  $L[\mathcal{B}] = w^{-1}\Sigma^*w$ . Let  $u \in \Sigma^*$  such that  $\lambda(v_0, u) = v_0$ , by (4) this is equivalent to  $\Sigma^*wu \subseteq \Sigma^*w$ , and it is not difficult to see that this is also equivalent to  $wu \wedge_s w = w$ . In the proof of Proposition 2 we have seen that  $wu \wedge_s w = w$  is equivalent to  $\xi(q_n, u) = q_n$ , i.e.  $u \in L[\mathcal{A}_w] = w^{-1}\Sigma^*w$ .

The first inclusion in the statement of the theorem  $\Sigma^*w\Sigma^* \subseteq \text{Syn}(\mathcal{B})$  is a consequence of Proposition 1. Let us prove the second inclusion. The following claim is of use.

*Claim.* For any  $I_j$  with  $j \in Q$  there is at least a  $\sigma_w$ -class  $J_{v_h}$  such that  $J_{v_h} \subseteq I_j$  for some  $v_h \in H$ .

*Proof.* Indeed, this property clearly holds for the left ideal  $I_i$  containing  $J_{v_0} = \Sigma^*w$ . Thus consider any  $I_j$  for  $j \neq i$ . Since  $I_j$  is a left ideal, for any  $u \in I_j$  we get  $I_iu \subseteq I_j$ . In particular we get  $J_{v_0}u \subseteq I_iu \subseteq I_j$ .  $\square$

Take any  $u \in \text{Syn}(\mathcal{B})$ , thus there exists some  $v_i \in T$  such that  $J_{v_k}u \subseteq J_{v_i}$  for all  $v_k \in T$ . Using the Claim we conclude that there exists some  $i \in Q$  such that  $I_ju \subseteq I_i$  for all  $j \in Q$ . Therefore,  $Iu \subseteq I_i$  and by condition ii) of Definition 1 we obtain  $u \in \text{Syn}(\mathcal{A})$ .

Let us prove the last statement of the theorem. Consider the map  $\varphi : H \rightarrow Q$  defined by  $\varphi(v_h) = j$  where  $j \in Q$  is the unique index such that  $J_{v_h} \subseteq I_j$ . We claim that  $\varphi : \mathcal{B} \rightarrow \mathcal{A}$  is a homomorphism. Indeed, take any  $h \in H, a \in \Sigma$ , and put  $t = \lambda(h, a), r = \varphi(t), q = \varphi(h)$ . Since  $J_h \subseteq I_q, J_ha \subseteq J_t \subseteq I_r$  and  $J_ha \subseteq I_qa$ , then  $J_ha \subseteq I_r \cap I_qa$ . Therefore, by the property of reset left regular decompositions we get  $I_qa \subseteq I_r$ , whence  $\varphi(\lambda(h, a)) = r = \delta(\varphi(h), a)$ , and this concludes the proof of the theorem.  $\square$

**Theorem 7.** *Cerny's conjecture holds if and only if for any  $\mathcal{B} \in \mathcal{L}(\Sigma)$  and  $\rho \in \text{Cong}_k(\mathcal{B})$  for all  $k < \sqrt{\|\text{Syn}(\mathcal{B})\|} + 1$  we have*

$$\|\text{Syn}(\mathcal{B}/\rho)\| < \|\text{Syn}(\mathcal{B})\|$$

*Proof.* Since Cerny's conjecture holds if and only if it holds for strongly connected automata, we can suppose without loss of generality that the automata considered are strongly connected. Thus, suppose that Cerny's conjecture holds for strongly connected synchronizing automata and let  $\mathcal{B} \in \mathcal{L}(\Sigma)$ ,  $\rho \in \text{Cong}_k(\mathcal{B})$  for some  $k < \sqrt{\|\text{Syn}(\mathcal{B})\|} + 1$ . Take  $I = \text{Syn}(\mathcal{B}/\rho)$ . By Proposition 1  $\mathcal{B}$  is strongly connected, thus a quotient automaton  $\mathcal{B}/\rho$  is strongly connected. Hence by Theorem 1 we have  $k \geq \text{rdc}(I) \geq \sqrt{\|I\|} + 1$ , i.e.  $\|\text{Syn}(\mathcal{B}/\rho)\| < \|\text{Syn}(\mathcal{B})\|$ .

Suppose that for any  $\mathcal{B} \in \mathcal{L}(\Sigma)$  and  $\rho \in \text{Cong}_k(\mathcal{B})$  for all  $k < \sqrt{\|\text{Syn}(\mathcal{B})\|} + 1$  the inequality in the statement of the theorem holds. Let  $\mathcal{A}$  be a strongly connected synchronizing automaton with  $k$  states. Let  $w$  be a reset word for  $\mathcal{A}$  of minimum length. For the word  $w$  we build the automaton  $\mathcal{B}$  of Theorem 6 associated to  $w$  such that  $\mathcal{A}$  is a homomorphic image of  $\mathcal{B}$ . Actually from the proof of Theorem 6 it follows that  $\mathcal{A}$  can be viewed as a quotient automaton  $\mathcal{B}/\rho$  for some  $\rho \in \text{Cong}_k(\mathcal{B})$ . By the same theorem we also have  $\Sigma^*w\Sigma^* \subseteq \text{Syn}(\mathcal{B}) \subseteq \text{Syn}(\mathcal{A})$ , hence

$$\|\text{Syn}(\mathcal{A})\| = |w| = \|\text{Syn}(\mathcal{B})\|.$$

Therefore, by the statement of the theorem we must have

$$k \geq \sqrt{\|\text{Syn}(\mathcal{B})\|} + 1 = \sqrt{\|\text{Syn}(\mathcal{A})\|} + 1$$

whence  $\mathcal{A}$  satisfies Cerny's conjecture.  $\square$

**Proposition 3.** *Let  $I = w^{-1}\Sigma^*w$  for some  $w \in \Sigma^*$ . The syntactic complexity of  $I$  is equal to*

$$\sigma(I) = |w| + 1 + |\text{Pref}(w)| + |\text{Fact}(w)| + |\text{Suff}(w)| - |\text{Pref}_{\text{syn}}(w)|.$$

*Proof.* Let  $\mathcal{A}_w = \langle P(w), \Sigma, \xi, q_n, \{q_n\} \rangle$  be the minimal automaton recognizing  $I$  as in Proposition 2. So  $P(w) = \{q_0, \dots, q_n\}$  is the set of prefixes of the word  $w$ ,  $|q_i| = i$  for all indices  $i$ , and  $\xi(q_i, a) = (q_i a) \wedge_s w$  for any  $q_i \in P(w)$ ,  $a \in \Sigma$ . By Proposition 1  $w$  is a reset word for  $\mathcal{A}_w$  and  $\mathcal{A}_w$  is strongly connected. Thus, since  $w \in I$ , we have  $P(w) \cdot w = \{q_n\}$ . Furthermore, for each  $q_i \in P(w)$  there exists some  $u \in \Sigma^*$  such that  $q_n \cdot u = q_i$ , hence  $P(w) \cdot wu = \{q_i\}$ . Note that  $|P(w)| = n + 1$ , so we can find  $n + 1$  reset words for  $\mathcal{A}_w$  defining pairwise different transformations of the automaton.

Take any  $u, v \in \text{Fact}(w)$ ,  $u \neq v$ . There exist some  $q_i, q_j \in P(w)$  such that  $q_i \cdot u = q_i u = q_{i+|u|}$  and  $q_j \cdot v = q_j v = q_{j+|v|}$  (see the illustration below).

$$\underbrace{\overbrace{w[1] \dots w[i]}^{q_i} \overbrace{w[i+1] \dots w[i+|u|]}^u}_{q_{i+|u|}} \quad \underbrace{\overbrace{w[1] \dots w[j]}^{q_j} \overbrace{w[j+1] \dots w[j+|v|]}^v}_{q_{j+|v|}}$$

Clearly,  $q_0 \cdot u <_p q_{i+|u|} = q_i \cdot u$  and  $q_0 \cdot v <_p q_{j+|v|} = q_j \cdot v$ , hence  $u$  and  $v$  are not reset words for  $\mathcal{A}_w$ . Without loss of generality suppose that  $|u| \leq |v|$ . We show that  $u$  and  $v$  define different transformations of  $\mathcal{A}_w$  by considering the following cases.

**Case 1.** Assume  $|u| < |v|$ . If  $i = j$  then  $q_i \cdot u = q_{i+|u|} \neq q_{i+|v|} = q_i \cdot v$ . If  $i < j$  then  $q_j \cdot u \neq q_j \cdot v$  since  $q_j \cdot u = q_k$  for some  $0 \leq k < n$  and  $q_k <_p q_j v = q_j \cdot v$ . If  $i > j$  then using an analogous argument we have  $q_j \cdot u \neq q_j \cdot v$ .

**Case 2.** Assume that  $|u| = |v|$ . If  $i = j$  then  $q_i \cdot u = q_i \cdot v$  since  $|u| = |v|$ . Thus  $u = v$ , which is a contradiction. If  $i < j$  then  $q_j \cdot u \neq q_j \cdot v$  since  $q_j \cdot u = q_k$  for some  $0 \leq k < n$  and  $q_k <_p q_j v = q_j \cdot v$ . If  $i > j$  then using the same an analogous argument we have  $q_i \cdot u \neq q_i \cdot v$ .

Take any suffixes  $s, t \in \text{Suff}(w)$ ,  $s \neq t$ . There exist some  $q_i, q_j \in P(w)$  such that  $q_i \cdot s = q_i s = w = q_n$  and  $q_j \cdot t = q_j t = w = q_n$ . Without loss of generality suppose that  $|s| \leq |t|$ . If  $|s| = |t|$  then  $s = t$ , which is a contradiction. So we may assume  $|s| < |t|$ . Thus,  $q_j \cdot s \neq q_j \cdot t$  since  $q_j \cdot s <_p w = q_j \cdot t$ . Therefore, different suffixes define different transformations of  $\mathcal{A}_w$ . Furthermore,  $q_0 \cdot s \leq_p s \neq w = q_i \cdot s$ , so  $s \notin \text{Syn}(\mathcal{A}_w)$ . Analogously,  $t \notin \text{Syn}(\mathcal{A}_w)$ . It remains to show that there is no proper suffix  $t$  defining the same transformation of  $\mathcal{A}_w$  as some inner factor different from  $t$ . Let  $u \in \text{Fact}(w)$ ,  $t \in \text{Suff}(w)$  and  $t \neq u$ . Again, consider  $q_i, q_j \in P(w)$  such that  $q_i \cdot u = q_i u$  and  $q_j \cdot t = w = q_n$ . If  $i \leq j$  then, since  $u \notin \text{Suff}(w)$ ,  $q_i \cdot u = q_i u <_p w = q_j \cdot t$ . If  $i > j$  then  $q_j <_p q_i$ , thus  $q_j \cdot u <_p q_i u = q_i \cdot u <_p w = q_j \cdot t$ . Hence, we get that  $u$  and  $t$  define different transformations of  $\mathcal{A}_w$ .

Take any prefixes  $x, y \in \text{Pref}(w)$ ,  $x \neq y$ . Since  $q_0 \cdot x = x \neq y = q_0 \cdot y$ , we get that  $x$  and  $y$  define different transformations of  $\mathcal{A}_w$ . We now show that proper prefixes define transformations which differ from transformations defined by any proper factor or a suffix. Indeed, take any two different words  $x \in \text{Pref}(w)$  and  $u \in \text{Fact}(w)$  such that  $u \neq x$ . If  $|x| \geq |u|$  then  $q_0 \cdot x = x \neq q_0 \cdot u$ . If  $|x| < |u|$  then  $q_i \cdot x \neq q_i \cdot u$ , where  $q_i \cdot u = q_i u$  for some  $q_i \in P(w)$ . Consider now any two different words  $x \in \text{Pref}(w)$  and  $t \in \text{Suff}(w)$  such that  $t \neq x$ . If  $|x| \geq |t|$  then  $q_0 \cdot x = x \neq q_0 \cdot t$  (otherwise  $x$  would be a suffix of  $w$ ). If  $|x| < |t|$  then  $q_j \cdot x \neq q_j \cdot t$ , where  $q_j \cdot t = w$  for some  $q_j \in P(w)$ . If there is some  $x \in \text{Pref}(w) \cap \text{Pref}_{\text{syn}}(w)$  then  $x \in \text{Syn}(\mathcal{A}_w)$ , but each reset word for  $\mathcal{A}_w$  belongs to the one of  $n + 1$  equivalence classes defined earlier. Therefore, we have proved

$$\sigma(I) \geq |w| + 1 + |\text{Pref}(w)| + |\text{Fact}(w)| + |\text{Suff}(w)| - |\text{Pref}_{\text{syn}}(w)|$$

Now, take any  $z \in \Sigma^*$  such that  $z \notin \text{Syn}(\mathcal{A}) \cup \text{Pref}(w) \cup \text{Fact}(w) \cup \text{Suff}(w)$ . It remains to show that  $z$  does not define a new transformation of  $\mathcal{A}_w$  that differs from transformations corresponding to reset words of  $\mathcal{A}_w$ , prefixes, suffixes or factors of  $w$ . Note that  $z \neq w$  since  $w \in \text{Syn}(\mathcal{A}_w)$ . If  $|z| \geq |w|$  then  $z \in \text{Syn}(\mathcal{A}_w)$  since by Lemma 1 we have  $q_i \cdot z = q_i z \wedge_s w = z \wedge_s w$  for all  $q_i \in P(w)$ , so we are done. Assume that  $|z| < |w|$ , and put  $q_0 \cdot z = q_k$ . The latter means that  $q_k$  is the maximal suffix of  $z$  which appears in  $w$  as a prefix. We may assume that  $q_i \cdot z = q_j \neq q_k$  for some  $q_i \in P(w)$ , otherwise  $z \in \text{Syn}(w)$ . By the definition of  $q_k$  we have  $q_k <_s q_j$ . Furthermore, by the definition of  $q_j$  we get  $q_j <_s q_i z$ . We have

two cases: either  $q_j \leq_s z$ , or  $z <_s q_j$ . If  $q_j \leq_s z$  then, since  $q_k$  is the maximal suffix of  $z$  which is also a prefix of  $w$ , we get  $q_j \leq_s q_k$ , a contradiction. In the second case we have  $z <_f q_j$  then, since  $q_j \leq_p w$ , we get  $z <_f w$ , a contradiction. So we have  $\sigma(I) = |w| + 1 + |\text{Pref}(w)| + |\text{Fact}(w)| + |\text{Suff}(w)| - |\text{Pref}_{syn}(w)|$ .  $\square$